INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A Bayesian Regularization Procedure for a Better Extremal Fit

Jean Diebolt — Myriam Garrido — Catherine Trottier

N° 4211

juin 2001

THÈME 4

rapport de recherche

ISRN INRIA/RR--4211--FR+ENG

ISSN 0249-6399

A Bayesian Regularization Procedure for a Better Extremal Fit

Jean Diebolt * , Myriam Garrido , Catherine Trottier †

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 4211 — juin 2001 — 21 pages

Abstract: In Structural Reliability, special attention is devoted to model distribution tails. This is important when one wants to estimate the occurrence probability of rare events as critical failures, extreme charges, resistance measures, frequency of stressing events, etc. People try to find distribution models having a good overall fit to the data. Particularly, the distributions are strongly required to fit the upper observations and provide a good picture of the tail above the maximal observation. Specific goodness-of-fit tests such as the ET test can be constructed to check this tail fit. Then what can we do with distributions having a good central fit and a bad extremal fit ? We propose a regularization procedure, that is to say a procedure which preserves the general form of the initial distribution and allows a better fit in the distribution tail. It is based on Bayesian tools and takes the opinion of experts into account. Predictive distributions are proposed as model distributions. They are obtained as a mixture of the model family density functions according to the posterior distribution. Therefore, they are rather smooth and can easily be simulated. We numerically investigate this method on normal, lognormal, exponential, gamma and Weibull distributions. Our method is illustrated on both simulated and real data sets.

Key-words: Goodness-of-fit Test, Tail distribution, Extreme test, Upper quantile, Mixture of distributions

* CNRS, Université de Marne-la-vallée, Analyse et Mathématiques Appliquées, 5 Bd Descartes, Champs-sur-Marne, 77454 Marne-la-vallée Cedex 2, France. `diebolt@math.univ-mlv.fr`

† Université Charles de Gaulle - Lille 3, Gremars, B.P. 149, 59653 Villeneuve d'Ascq Cedex, France. `trottier@univ-lille3.fr`

Une procédure de régularisation bayérienne pour une meilleure adéquation extrême

Résumé : En fiabilité structurale, une attention particulière est portée aux queues de distribution. En effet, on cherche souvent à évaluer la probabilité d'occurrence d'événements rares, tels que charges extrêmes, défaillance de dispositifs critiques, mesures de ténacité, fréquences de stress impliquant un vieillissement accéléré, etc. Étant donné un échantillon, il s'agit de déterminer des modèles de lois de probabilité qui s'ajustent globalement aux données. En particulier, les lois cherchées doivent correctement représenter et modéliser les plus grandes observations, et donner une bonne estimation de la queue de distribution au-delà de l'observation maximale. Des test d'adéquation spécifiques, comme le test ET, peuvent être construits pour vérifier la qualité de l'ajustement en queue de distribution. Mais que faire lorsque les lois testées révèlent une bonne adéquation centrale, et une mauvaise adéquation extrême ? Nous proposons ici une procédure de régularisation, c'est-à-dire une procédure qui conserve l'allure générale de la loi initiale et permet un meilleur ajustement de la queue de distribution. Cette procédure est basée sur la mise en œuvre d'outils bayésiens, et prend en compte l'opinion d'experts. Les lois prédictives qui en découlent sont proposées comme nouveaux modèles. Elles sont obtenues comme mélange continu de densités du modèle initial par rapport à la loi a posteriori. Elles sont donc relativement lisses et aisément simulables. Nous détaillons numériquement cette méthode pour les modèles normal, lognormal, exponentiel, gamma et Weibull. Nous illustrons la méthode sur des données simulées et des données réelles.

Mots-clés : test d'adéquation, queue de distribution, test extrême, quantile extrême, distributions de mélange.

1 Introduction

The present paper lies in the context of tail approximation and extreme quantile estimation.

Part of the reliability and safety engineering activity is related to the control of extreme values (see, e.g., the monograph Castillo 1988). These values may be quantities entering complex physical systems or outputs of some components of it. In any case a good stochastic modelization is sought (see numerical experiments in Hughey 1991 or Moon, Lall, and Bosworth 1993). In general, parametric model families are selected through usual goodness-of-fit tests (see, e.g., D'Agostino and Stephens 1986). However, such tests do not focus on the tail behaviour of the tested distributions. Only the central behaviour is really checked. This may prove to be very dangerous when working afterwards on distribution tails for upper quantile estimation, rare event prediction, extreme value simulation, *etc.*. An illustration of that is given by Ditlevsen (1994) (see also Hahn and Meeker 1982) and the consequences of selecting a wrong model with a fair central fit are also investigated by Diebolt, Durbec, El Aroui, and Villain (2000). In this paper, our goal is to construct smooth and workable parametric models with a good overall fit for both the central and extreme ranges of the data set. We wish that the distributions in these models have smooth density functions. Moreover, these distributions should be easily simulated and allow for analytical calculations or simple numerical computations. For instance, we have to be able to compute the distribution and the quantile functions. To check tail goodness-of-fit, extreme tests have recently been investigated. They are naturally linked to techniques for upper quantile estimation (see, e.g., Diebolt et al. 2000 and the monograph Embrechts, Klüpperberg, and Mikosch 1997). In this work, we will use the so-called ET (Exponential Tail) test. It is based on an exponential approximation of the distribution of excesses-over-a-threshold when the model belongs to Gumbel's maximum domain of attraction. See, e.g., Pickands (1975), Breiman, Stone, and Kooperberg (1990), and Girard and Diebolt (1999). We sketch the main features of the ET test in section 2. More details can be found in Garrido and Diebolt (2000), and in Diebolt, Garrido, and Girard (2001).

We thus have in hands two kinds of tests: one for the central fit and the other one for the tail fit. In practice, when testing the different distributional hypotheses from a set of models, we may end up with none of them being accepted both centrally and for the tail. In other words, a parametric model providing a good central description of the observations is not necessarily well suited to the largest observations (Ditlevsen 1994). In order to build a distribution with a good overall fit, we propose a *regularization procedure* in this paper. Since we do not assume any structural explanation for the tail bad fit, the starting point of this procedure is a centrally accepted distribution and all our attention is then focused on producing a sufficient tail distortion to end with a satisfying tail description. This is the main idea of our regularization method.

Another requirement of the present work is to take advantage of earlier information. Indeed, we assume that we have expert opinions. The latter may be subjective or related to previous experimental results. In this context, our regularization procedure follows a Bayesian strategy which is described in section 3. Other Bayesian approaches to extreme quantile estimation can be found, e.g., in Coles and Dixon (1999), Coles and Powell (1996), and Coles and Tawn (1999). The Bayesian tools not only enable us to take into account this kind of information but also to get a smooth distortion of the tail, which is one of our requirements. The main output of the Bayesian procedure described in subsection 3.1 is the predictive distribution, which is then proposed as the new model distribution. In subsection 3.2, predictive distributions are given for standard models: exponential, gamma, normal, lognormal and Weibull. In this framework there are still questions about the definition of prior information and the way it can enter the model. Subsection 3.3 defines the form of expert opinions that we can dispose and their use to determine hyperparameters.

Section 4 provides numerical results on our Bayesian regularization procedure, for both simulated and real data sets.

2 The ET Test.

Before setting up any regularization procedure, the first step is to check whether a parametric model $\{F_\theta \ : \theta \in \Theta\}$ provides an acceptable approximation to the distribution of the sample $X_1, \dots, X_n$, both centrally and in the upper tail. Actually, we are mainly interested in extrapolating the tail of the distribution above the maximal observation $X_{(n)}$.

Let us assume that at reasonable significance levels, usual goodness-of-fit tests (e.g., Cramér-von Mises or Anderson-Darling, see D'Agostino and Stephens 1986) have not rejected the null hypothesis H_0 that F (the distribution function of the X_i's) has the form F_θ. As already mentioned, such procedures test the adequacy of the model to the whole observed sample $\underline{x}_n = (x_1, \dots, x_n)$, hence essentially test the central part of the distribution. Now, with the help of the ET test, we focus on tail behaviour.

The ET test is roughly based on the comparison between two different estimates of an extreme quantile of order $1-p$, the nonparametric estimate $\widehat{q}_{ET}$ and the parametric estimate under H_0, $\widehat{q}_{\text{param}} = F^{-1}_{\widehat{\theta}_n}(1-p)$, where $\widehat{\theta}_n$ is the maximum likelihood estimate of θ.

We suppose that F and F_θ are in Gumbel's maximum domain of attraction, DA(Gumbel) (e.g., Castillo 1988). This domain contains the most usual models, e.g., normal, lognormal, exponential, Weibull and gamma distributions that we consider here.

2.1 The ET estimation of upper quantiles (nonparametric estimation).

The excesses above a threshold u are the $X_i - u$'s, $1 \le i \le n$, where $X_i > u$. Their common distribution function is

$$F_u(y) = P\left(X - u \le y \,|X > u\right) = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad \forall\, y \ge 0. \tag{1}$$

The theorem of Pickands (1975) implies that if $F \in \mathrm{DA(Gumbel)}$ and the upper endpoint $\omega(F) = \inf\{x : F(x) = 1\}$ is infinite (which holds for the models we consider), then

$$\lim_{u \to \infty} \sup_{0 \le y < \infty} \left| F_u(y) - \left(1 - \exp\left(-\frac{y}{\sigma(u)}\right)\right)\right| = 0, \tag{2}$$

where $\sigma(u)$ is some function of u defined up to asymptotic equivalence as $u \to \infty$.

To apply this, denoting by $x_{(1)} \le \ldots \le x_{(n)}$ the ordered observed sample, we take a threshold of the form $u_n = x_{(n-k_n)}$, where $k_n \to \infty$ and $\lim_{n\to\infty} k_n/n = 0$. Let us denote by $y_1, \ldots, y_{k_n}$ the corresponding k_n excesses. Under H_0 and for n large enough, we can roughly consider that these excesses are independent and exponentially distributed. Therefore, we can approximate $F_{u_n}(y)$ by $1 - \exp(-y/\widehat{\sigma}_n)$, where $\widehat{\sigma}_n = k_n^{-1}\sum_{j=1}^{k_n} y_j$ is the empirical mean of the excesses. Using (1) and the exponential approximation (2), we obtain the ET (*Exponential Tail*) estimate of the quantile of order $1-p$,

$$\widehat{q}_{ET} = x_{(n-k_n)} + \widehat{\sigma}_n \ln\left(\frac{k_n}{np}\right). \tag{3}$$

2.2 The ET test based on parametric bootstrap.

We first construct a confidence interval under H_0 for the approximation error between the true quantile and its ET approximation. To this end, we approximate the sampling variations of $\widehat{q}_{\text{param}}$ and $\widehat{q}_{ET}$ through parametric bootstrap. In order to generate a sample of values of $\widehat{q}_{ET}$ and $\widehat{q}_{\text{param}}$, we first independently generate N iid samples of size n from $F_{\widehat{\theta}_n}$. Then, for the jth bootstrap sample, we compute the estimates $\widehat{q}_{ET}^{*(j)}$ and $\widehat{q}_{\text{param}}^{*(j)} = F^{-1}_{\widehat{\theta}_n^{*(j)}}(1-p)$. This yields a sample of N values $\widehat{\delta}^{*(j)} = \widehat{q}_{\text{param}}^{*(j)} - \widehat{q}_{ET}^{*(j)}$ of the approximation error. Based on this simulated sample we can compute empirical confidence intervals $IC_{\delta,PB}$ for the approximation error δ.

ET-PB test (complete version based on parametric bootstrap): At significance level α, we do not reject H_0 if $\widehat{\delta} = \widehat{q}_{\text{param}} - \widehat{q}_{ET}$, the approximation error computed from the original sample, lies within the $(1-\alpha)$–confidence interval $IC_{\delta,PB}$.

Simplified version of the test: If we consider that the sampling variations of $\widehat{q}_{\text{param}}$ are small with respect to those of $\widehat{q}_{ET}$, then we only have to bootstrap $\widehat{q}_{ET}$ and compute confidence intervals $IC_{ET,PB}$ for q_{ET}. Therefore, at significance level α, we do not reject H_0 if $\widehat{q}_{ET}$ lies within the $(1-\alpha)$–confidence interval $IC_{ET,PB}$.

This simplified version of the ET test is basically of interest in two cases. First, when the computation of the maximum likelihood estimate is not easy. Second, when the computation of the quantile function F_θ^{-1} is hard. For instance, such difficulties arise for mixture distributions and for the distributions built with our regularization procedure.

3 A Bayesian regularization approach

Let us now consider that $F_{\widehat{\theta}_n}$ is a distribution accepted centrally but not in the upper tail. Our aim is to build a model with a good overall fit based on $F_{\widehat{\theta}_n}$. This new model should fit the largest observations and take into account information given by experts.

A first idea is to aggregate $F_{\widehat{\theta}_n}$ up to some threshold with the exponential tail approximation of F. Unfortunately, this leads to a discontinuous density function. Moreover, this method cannot enter prior information.

3.1 A Bayesian procedure

Transforming the opinions of experts into prior information, we present a Bayesian procedure which is a smooth regularization method. It consists in relaxing the condition $\theta = \widehat{\theta}_n$ by putting a prior distribution Π_γ on θ, with hyperparameters γ. In subsection 3.3, we explain how this prior distribution takes into account information given either by experts or by the nonparametric ET estimate of the tail. This enables us to attract θ towards a range of values more suited to the shape of the tail.
The posterior distribution is given by

$$\forall \theta \in \Theta \qquad \Pi_\gamma(\theta|\underline{x}_n) \propto (\Pi_{i=1}^n f_\theta(x_i))\ \Pi_\gamma(\theta) \ \ .$$

We propose to take the so-called *predictive* density function as the regularized density function. This predictive density function is obtained by integrating $f_\theta(x)$ with respect to the posterior density of θ:

$$\forall x \in I\!R \qquad f_\gamma(x|\underline{x}_n) = \int_\Theta f_\theta(x)\Pi_\gamma(\theta|\underline{x}_n)\, d\theta \ \ .$$

It is an infinite mixture with respect to the posterior distribution of θ and can easily be simulated.

3.2 In practice

In this paper, we are interested in the following five standard distribution families: exponential, gamma, normal, lognormal and Weibull. All these distributions, except the exponential one, involve a two-dimensional parameter θ. To facilitate calculations, we have chosen to put a prior distribution on only one of the two components of θ, the other one being kept constant at its estimated value. As far as possible, we put a prior on the parameter component which appears to be the most sensitive regarding to tail behaviour. In the normal case, we take $1/\sigma^2$. For $\mathcal{G}amma(a, b)$ distributions we take the scale parameter b, since the exponential is the most influent term for tail behaviour. For Weibull $\mathcal{W}(\alpha,\, \beta)$ distributions, although the shape parameter β is the most important for the tail, we first put a prior on the scale parameter α to avoid heavy computations. The case of a prior on the shape parameter is postponed to subsection 3.4. The treatment of both lognormal and Weibull (with fixed

$\beta = \widehat{\beta}_n$) distributions reduces to that of the normal and exponential distributions, respectively, through a transformation of the variables. We use conjugate priors whenever it is possible. In each situation studied in this subsection (normal with fixed mean, exponential and gamma with fixed shape parameter) the prior is a conjugate gamma distribution. This enables us to write the predictive density functions in closed form. Table 1 displays the predictive models (see the appendix for the parametrisations we use). As can be seen, we obtain:

- for the exponential case, a generalized Pareto distribution,
- for the gamma case, a rescaled beta(II) distribution,
- for the normal case, a density function which cannot be expressed in terms of documented distributions and can be seen as an extended version of a Student density. In that case, numerical integration is required to compute quantiles. In contrast, simulation from this predictive density is easy. Since it is a mixture of normal distributions $\mathcal{N}(\mu,\, 1/\theta)$ with mixing measure $\mathcal{G}amma(a',\, b')$ (see Table 1), we simply simulate $\theta \sim \mathcal{G}amma(a',\, b')$ and $X \sim \mathcal{N}(\mu,\, 1/\theta)$.

3.3 From expert opinions to prior distributions

Modelling the prior information, i.e. determining the hyperparameters of our gamma prior distribution, is an important point in the procedure. The regularized function depends on the unknown hyperparameters γ. Since we do not take a hierarchical Bayesian approach here, these parameters have to be specified.

To this end, we first define a variation interval $[\theta_1,\, \theta_2]$ for θ, which is given a confidence degree of $1-\varepsilon$ for some small $\varepsilon > 0$. Approaching the gamma prior distribution with a normal distribution centered on $[\theta_1,\, \theta_2]$, γ can be explicitly calculated.
To compute θ_1 and θ_2, we have to consider the two following situations separately.

1. With expert opinions
   The kind of information that can be expected from an expert (or a group of them, aggregating their opinions) concerns extreme values, see for instance Coles and Tawn (1999). We can ask the expert a value $q_{\max}$ of the quantity of interest that he thinks can be reached rarely. Then we have to quantify this rarity with his help. This is done by determining an upper bound p_1 and a lower bound p_2 for the risk associated to $q_{\max}$, i.e. the probability of overpassing $q_{\max}$. The bounds p_1 and p_2 can be of different orders. Typically, for a sample of size 50 to 100, we could take $p_1 = 10^{-2}$, and $p_2 = 10^{-4}$ or 10^{-5}. The effects of such possible choices are apparent in the numerical results in section 4. Finally, to compute the bounds θ_1 and θ_2, we interpret $q_{\max}$ as a quantile of orders $1-p_1$ and $1-p_2$, respectively:
   $$F_{\theta_i}^{-1}(q_{\max}) = 1 - p_i \qquad i = 1, 2.$$
   Notice that these values of θ_1 and θ_2 can be found analytically (exponential and associated Weibull cases) or numerically (normal, gamma and lognormal cases).

$f_\theta(x)$	$\Pi_\gamma(\theta)$	$\Pi_\gamma(\theta \mid \underline{x}_n)$	$f_\gamma(x \mid \underline{x}_n)$
$\mathcal{E}xp(\theta)$	$\mathcal{G}amma(a,b)$	$\mathcal{G}amma(a',b')$ $\text{with} \quad a' = a+n$ $b' = b + \sum_{i=1}^{n} x_i$	$GPD(\sigma,k)$ $\text{with} \quad k = -(a+n)^{-1}$ $\sigma = \frac{n\overline{x}+b}{n+a}$
$\mathcal{G}amma(\alpha,\theta)$	$\mathcal{G}amma(a,b)$	$\mathcal{G}amma(a',b')$ $\text{with} \quad a' = a+n\alpha$ $b' = b + \sum_{i=1}^{n} x_i$	$\frac{\Gamma(\alpha+a')}{\Gamma(\alpha)\Gamma(a')b'}(\frac{x}{b'})^{\alpha-1}(1+\frac{x}{b'})^{-(\alpha+a')}$
$\mathcal{N}(\mu,\frac{1}{\theta})$	$\mathcal{G}amma(a,b)$	$\mathcal{G}amma(a',b')$ $\text{with} \quad a' = a+n/2$ $b' = b + \sum_{i=1}^{n} (x_i-\mu)^2/2$	$\frac{\Gamma(a'+1/2)}{\sqrt{(2\pi b')}\Gamma(a')}(1+\frac{(x-\mu)^2}{2b'})^{-(a'+1/2)}$
$X \sim \mathcal{LN}(\mu,\frac{1}{\theta})$ $Y = \ln(X) \sim \mathcal{N}(\mu,\frac{1}{\theta})$	Use Normal case for Y and with sample $\underline{y}_n = \ln(\underline{x}_n)$		
$X \sim \mathcal{W}(\alpha,\beta)$ $Y = X^\beta \sim \mathcal{E}xp(\theta)$	Use Exponential case for Y and with sample $\underline{y}_n = (\underline{x}_n)^\beta$ with $\theta = 1/\alpha^\beta$		

Table 1: Distributions: model, prior, posterior and predictive.

2. Without expert opinions
   Let $\theta_{ET} = \lim_{p\to 0} \theta_p$, where θ_p is defined by $q_{ET}(p) = F_{\theta_p}^{-1}(1-p)$. This means that the $(1-p)$-quantile of F_{θ_p} coincides with the ET approximation, $q_{ET}(p)$, of the true quantile q_{1-p}. We obtain $\theta_{ET} = 1/\widehat{\sigma}_n$ for exponential, gamma and Weibull (with $\widehat{\sigma}_n$ computed from the transformed sample) distributions, and $\theta_{ET} = 0$ for normal and lognormal distributions. The bounds θ_1 and θ_2 are now chosen to allow for a compromise between (after a possible index permutation) $\theta_1 = \widehat{\theta}_n$ and $\theta_2 = \theta_{ET}$, where $\widehat{\theta}_n$ and θ_{ET} represent the central and tail behaviours, respectively.

Case 1 corresponds to a true Bayesian point of view, where we take prior information given by experts into account. Case 2 is rather an artificial Bayesian detour to include the information contained in the two possible estimates of θ.

3.4 The shape parameter of the Weibull distribution

In this subsection, we focus on priors for the shape parameter β, the scale parameter $\alpha = \widehat{\alpha}_n$ being kept constant.

Unfortunately, there is no conjugate prior in this case. In order to choose a suitable prior family, we first remark that the shapes of Weibull densities are very different whether $0 < \beta < 1$ or $\beta > 1$. Actually, a predictive density mixing Weibull densities with $0 < \beta \leq 1$ for some of them and $\beta > 1$ for the others would not fit the sample since the estimated Weibull distribution, $\mathcal{W}(\widehat{\alpha}_n, \widehat{\beta}_n)$, is assumed to be centrally accepted, and either $0 < \widehat{\beta}_n \leq 1$ or $\widehat{\beta}_n > 1$ (but not both).

Hence, we first require that the support of the prior distributions is contained either in $[0, 1]$ or in $[1, \infty)$. Therefore, we choose a family of priors with a compact support $[\ell_m, \ell_M]$ included either in $[0, 1]$ or in $[1, \infty)$. This interval must be chosen by the user at the beginning of the procedure. The value of $\widehat{\beta}_n$ can help to locate ℓ_m and ℓ_M. Notice that in Reliability, β cannot be much larger than 10.

Since the interval $[\theta_1, \theta_2]$ determined from expert opinions (see 3.3) must be contained into $[\ell_m, \ell_M]$, we cannot accept that $\theta_1 < 1 < \theta_2$. If such a case happens, there is a conflict between the central model and the prior information on the tail. The same contradiction appears when the interval $[\theta_1, \theta_2]$ is on one side of 1 and $\widehat{\theta}_n$ is on the other side. In both cases, the procedure must be stopped.

As the priors are compactly supported, a natural choice is the family of beta distributions with parameters $\gamma = (a, b)$ (see the appendix). With these priors, the posterior density $\Pi_\gamma(\theta|\underline{x}_n)$ has no closed form because the normalizing factor cannot be specified analytically. Hence, the predictive density has no closed form either. Given the sample $\underline{x}_n$, for each value of ℓ_m, ℓ_M, a and b, intensive numerical methods are necessary to compute the predictive distribution and extreme quantiles.

To specify these hyperparameters a and b, we first have to calculate θ_1 and θ_2 as in subsection 3.3, with $\theta_{ET} = 1$. Then, we consider three cases depending on the confidence we have in $[\theta_1,\, \theta_2]$.

1. If our confidence concerning $[\theta_1,\, \theta_2]$ is high, then we approximate the beta distribution by a normal distribution centered on $[\theta_1,\, \theta_2]$, with the same mean and variance. We then interpret $[\theta_1,\, \theta_2]$ as a $(1-\varepsilon)$ confidence interval. As can be seen in the left panel of Figure 1, only a small fraction of the distribution is outside $[\theta_1,\, \theta_2]$ in that case.

2. If we have a medium confidence in $[\theta_1,\, \theta_2]$, we want the beta distribution to be nonzero on an interval significantly larger than $[\theta_1,\, \theta_2]$. Preliminary numerical experiments have shown that when both parameters are not far from 5, we obtain a bell-shaped distribution putting significant weight on the major part of the domain $[\ell_m,\, \ell_M]$. Then, we propose to take one of the parameters equal to 5 and to take the mode of the distribution at the middle of $[\theta_1,\, \theta_2]$. If the mode is larger than the middle of $[\ell_m,\, \ell_M]$, we recommend to choose $a = 5$, and to take $b = 5$ in the other case. This situation is illustrated by the central panel of Figure 1.

3. If we have a low confidence in $[\theta_1,\, \theta_2]$, we want that the beta distribution has roughly the shape of a uniform distribution (obtained for $a = b = 1$). As above, experiments lead us to set one parameter equal to 1.2 and to locate the mode at the middle of $[\theta_1,\, \theta_2]$ to obtain an arch-shaped distribution (see the right panel of Figure 1). Again, we recommend to take $a = 1.2$ when the mode is larger than the middle of $[\ell_m,\, \ell_M]$, and $b = 1.2$ either.

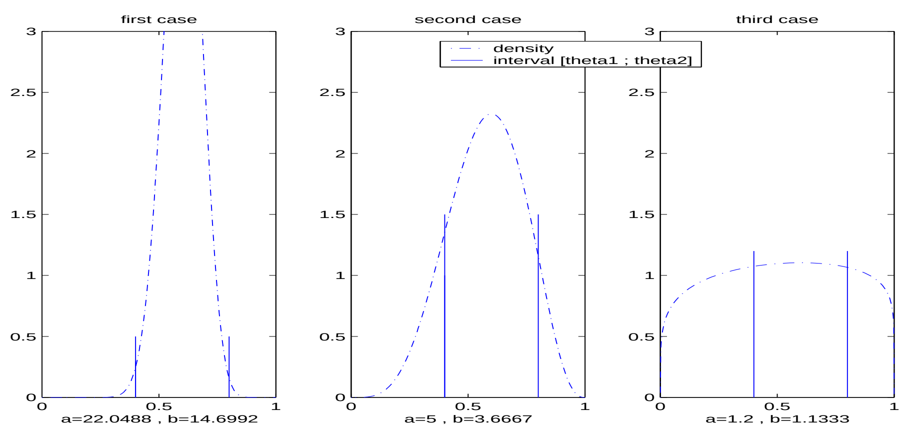

Figure 1: beta density for $\theta_1 = 0.4$, $\theta_2 = 0.8$ and $[l_m,\, l_M] = [0,\, 1]$

As already mentioned, we are interested in simulating from the predictive density. As in subsection 3.2, this reduces to simulating from the posterior density. Let $L_n(\theta) =$

$\prod_{i=1}^{n} f_\theta(x_i)$ be the likelihood function and $L_n(\widehat{\theta}_n)$ be its maximum value. We have implemented the following acceptation/rejection algorithm:

1. `Simulate` $Y \sim \mathcal{B}eta(a,\, b,\, [\ell_m,\, \ell_M])$`;`

2. `Simulate independently` $U \sim \mathcal{U}[0,\, 1]$`;`

   - `If` $U \leq \dfrac{L_n(Y)}{L_n(\widehat{\theta}_n)}\Pi_\gamma(Y)$ `then take` $\theta = Y$`;`
   - `Otherwise reject` Y`.`

The mean number of rejections before an acceptation occurs is the ratio $L_n(\widehat{\theta}_n) / \int_{\ell_m}^{\ell_M} L_n(\theta)\Pi_\gamma(\theta)\, d\theta$. In some experiments, we have found a ratio of the order of 100, but in general it is smaller.

Putting a prior on the shape parameter β of a Weibull distribution is much more complicated than putting a prior on the scale parameter α. However, it can help detect situations where the central and tail behaviours are conflictual.

Performances of this approach and the other ones are investigated and compared in section 4.

4 Numerical results

4.1 Simulation results

We first consider the simple cases (with no heavy computational difficulty) from the set of normal, exponential, gamma and lognormal distributions. We will then consider Weibull distributions for which the case of the shape parameter is much more computationally demanding.

For a given model in our set of distributions, the starting point of these simulations is to exhibit a sample centrally accepted by the Cramér-von Mises test but rejected in the upper tail by the ET test. To this end, we choose a simulation distribution looking like the given model distribution in its central part but with a different tail. With such a sample, we estimated the model parameters and run the regularization procedure. The outputs are:

- Extreme quantiles estimated with the true distribution, the true model with estimated parameters, the estimated hypothesized model distribution, the predictive distribution and the ET method;

- The changes in the Cramér-von Mises distance – CVM distance – and in the ET-bootstrap confidence interval. Notice that the Cramér-von Mises test cannot be directly used for predictive distributions, since there are no referenced critical values. Also, we have used the simplified version of the ET test even for the initial models in order to compare with predictive distributions.

- A plot of the different survival functions, starting near the threshold.

For the prior information, we took as $q_{\max}$ the value of the true quantile from the simulated distribution of order $1-10^{-3}$. As concerned the ET test, the number of excesses k_n we took in practice was suggested by preliminary numerical results.

The exponential case

The simulation distribution used was a $\mathcal{W}(3,\ 1.3)$. We generated a sample of size 100. The estimated model was $\mathcal{E}xp(0.3710)$ with a CVM distance of 0.1789, the critical value for the corresponding 5%-test being 0.2216. The simplified ET test executed at the order $1-10^{-3}$ and at level 5 % provided the interval $IC_{ET,\,PB}=[12.2673\,,\,27.0092]$ for the ET quantile, whereas the estimate of q_{ET} for this sample, considering $k_n = 14$ excesses, was $\widehat{q}_{ET} = 12.2601$. So, in this case the exponential model was centrally accepted but rejected in the tail. Table 2 presents the results of the regularization procedure.

	predictive distribution			estim. model distri-bution	estim. true distri-bution	simul. true distri-bution	ET
	with expert opinion		without expert opinion				
	$p_1=10^{-2}$ $p_2=10^{-4}$	$p_1=10^{-3}$ $p_2=10^{-5}$					
$q_{0.99}$	11.2397	9.5427	11.5121	12.4124	9.9728	9.7120	9.0307
$q_{0.999}$	16.9822	14.3947	17.4131	18.6185	13.8661	13.2668	12.2601
$q_{0.9999}$	22.8082	19.3011	23.4131	24.8247	17.5191	16.5528	15.4895
θ_1	0.3471	0.5207	0.3710	$\times$			
θ_2	0.6942	0.8678	0.7130	$\times$			
d_{CVM}	0.4785	1.2683	0.4026	0.1789	(critical value : 0.2216)		
ET test	acc.	acc.	acc.	rej.			
IC_{ET}	11.5512 24.0674	9.2169 19.5390	11.5068 24.8141	12.2673 27.0092			

Table 2: Regularization results – exponential model – $n = 100$

The first part of this table highlights the relevance of this procedure, since the accepted exponential model produced overestimates quantiles (even more overestimated as the quantile order is higher), whereas whatever the kind of regularization used, the quantile estimated by the predictive distribution tends to be attracted towards the true value. This can also be seen on the plot of the survival functions in Figure 2, where the regularized distribution tail lies between the true and the model distribution tails.

As expected, the correction produced by the procedure is very soft. There are no big jumps neither in the quantile estimation nor in the interval IC_{ET}. Without expert opinions, the model parameter θ, initially estimated at 0.3710, is forced to vary with a probability of 0.95 within the interval $[0.3710,\quad 0.7130]$ (compromise between $\widehat{\theta_n}$ and θ_{ET}).

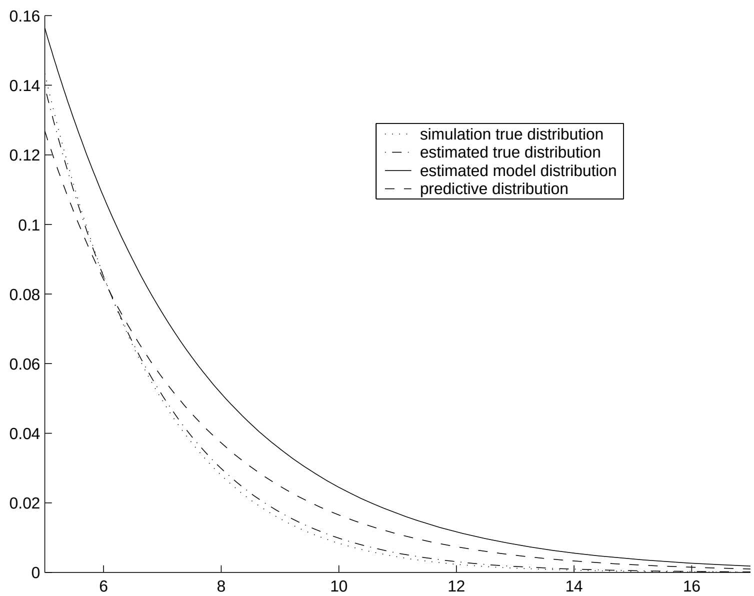

Figure 2: Exponential model – survival functions (starting from the threshold) – predictive distribution obtained with expert opinions ($p_1 = 10^{-2}$, $p_2 = 10^{-4}$)

This interval changes according to prior information. With more constraints on the tail ($p_1 = 10^{-3}$, $p_2 = 10^{-5}$), the regularized distribution has a better tail behaviour but its central part is less adequate ($d_{CVM} = 1.2683$). Without documented critical value for the predictive distribution, we cannot accept or reject the central goodness-of-fit of the regularized distribution.
In this case, we have succeeded in building a regularized distribution accepted in the tail, without an important deterioration of the CVM distance.

The normal case

The Student distribution with 4 degrees of freedom was used to simulate a sample of size 50. The estimated model was $\mathcal{N}(-0.0239, 1.5160)$ with a CVM distance of 0.1005 for a 5 % critical value of 0.1248. The confidence interval for the simplified ET test, computed at the order $1 - 10^{-3}$ with $k_n = 8$ excesses, was $IC_{ET,PB} = [2.7378\,, 7.0153]$ and the ET estimate was $\widehat{q}_{ET} = 8.4589$. Table 3 displays the results of the regularization procedure.

Unlike the previous case, here the true distribution has a heavier tail than the hypothesized model distribution and the estimation of q_{ET} on the simulated sample is larger than the upper limit of the confidence interval obtained by parametric bootstrap. Again, the regularization procedure succeeds in attracting the tail distribution towards better values.

	predictive distribution			estim. model distribution	simul. true distribution	ET
	with expert opinions		without expert opinion			
	$p_1 = 10^{-2}$ $p_2 = 10^{-4}$	$p_1 = 10^{-3}$ $p_2 = 10^{-5}$				
$q_{0.98}$	4.1666	3.7555	3.3421	3.0896	2.9985	4.0484
$q_{0.999}$	6.3629	5.7182	5.1771	4.6609	7.1732	8.4589
$q_{0.9999}$	7.8229	7.0920	6.5062	5.6141	13.0337	11.8489
θ_1	0.1045	0.1844	0	$\times$		
θ_2	0.2670	0.3512	0.4351	$\times$		
d_{CVM}	0.3249	0.2246	0.1313	0.1005	(critical value : 0.1248)	
ET test	acc.	acc.	acc.	rej.		
IC_{ET}	4.5788 11.3167	4.2197 10.0413	3.6643 8.8950	2.7378 7.0153		

Table 3: Regularization results – normal model – $n = 50$

As above, the correction is slight and the more constraints we put on the tail, the less the model is centrally adequate. There is no particular effect due to a smaller sample.

The Weibull case

We look at both the shape and scale parameters of the Weibull distribution. The gamma $\mathcal{G}amma(3, 3)$ distribution is used for simulations. The estimated model is $\mathcal{W}(1.1325, 1.5246)$ with a CVM distance of 0.0596 for a 5 % critical value of 0.1216. Applying both versions of the ET test at order $1 - 10^{-3}$ and for $k_n = 20$ excesses results in the interval $IC_{ET,\,PB}$ = [3.3442, 5.6401] for the ET quantile, $\widehat{q}_{ET} = 5.3416$, and $IC_{\delta,\,PB} = [-0.3852, 1.2964]$ for the approximation error, with $\widehat{\delta} = 1.3184$. The Weibull model is centrally accepted but rejected by the complete ET test, although not rejected by the simplified version of the test. Table 4 displays the detailed results. The Weibull parameters are estimated by maximum likelihood. The simulated expert provides the $1-10^{-3}$ quantile of the simulated true gamma distribution.

Again, the correction is soft. No big jumps appear neither in the quantile estimation nor in the Cramér-von Mises distance. Note that regularization does not always imply central deterioration. In this case for example, the CVM distance decreases for three of the computed predictive distributions. But note also that here the curve of the Weibull survival function (model hypothesis) is between the curve of the true survival function (gamma distribution) with simulated parameters and the one with estimated parameters.

	predictive distribution				estim.	estim.	simul.	ET
	shape parameter			scale	model	true	true	
	case 1	case 2	case 3	param.	dist.	dist.	dist.	
$q_{0.99}$	3.0700	3.1110	3.1230	3.0330	3.0837	3.3055	2.8020	3.6181
$q_{0.999}$	4.0400	4.1430	4.1710	3.9759	4.0232	4.5742	3.7430	5.3416
$q_{0.9999}$	4.9660	5.1770	5.2360	4.8245	4.8587	5.8052	4.6427	7.0650
θ_1	1.2775	1.2775	1.2775	0.6156	$\times$			
θ_2	1.8573	1.8573	1.8573	1.2312	$\times$			
d_{CVM}	0.0556	0.0585	0.0596	0.0536	0.0596	(critical value : 0.1216)		
ET test	acc.	acc.	acc.	acc.	acc.			
IC_{ET}	3.4471 5.6906	3.4743 5.8036	3.3876 5.8373	3.3709 5.5288	3.3442 5.6401			

Table 4: Regularization results – Weibull model – $n = 100$

4.2 Real data set results

The $n = 11$ data are welding defect heights provided by the French electricity company:

$$X = [1.80, 2.20, 2.50, 2.60, 2.20, 1.50, 1.70, 2.30, 2.20, 2.50, 1.30].$$

Applying the Anderson-Darling or the Cramér-von Mises test to this data set, only the exponential distribution was rejected at usual significance levels (i.e., $\alpha = 0.01, 0.05$ or 0.1). The complete version of the ET-BP test rejected the same distribution with $k_n = 4$ excesses, $p = 1/n$, 0.05 or 0.01, at significance level $\alpha = 0.05$.

For Bayesian regularization, the expert provided the value $q_{\max} = 3.2$mm and the following interpretations in terms of quantiles: either $p_1 = 10^{-2}$ and $p_2 = 10^{-3}$, or $p_1 = 10^{-2}$ and $p_2 = 10^{-4}$. Applying the Bayesian procedure to the normal model, we obtained the results displayed in Table 5.
We can see that the quantiles of both predictive distributions are very close to those of the normal distribution. Moreover, the Cramér-von Mises distance is almost the same for the model and the predictive distributions. We deduce that the regularization procedure keeps the predictive distributions close to the model. Therefore, the normal distribution seems to be in harmony with the data and the expert.

We then applied the regularization procedure to the lognormal model. The results are presented in Table 6.
The quantiles obtained for both predictive distributions are far from those of the lognormal distribution and distinctly get closer to those of the normal model. This confirms the accordance between the normal model and the expert. Figure 3 shows the densities of the normal, lognormal and predictive (based on the lognormal model, with $p_1 = 10^{-2}$ and $p_2 = 10^{-4}$) distributions.

		predictive distribution $p_1 = 10^{-2}$ $p_2 = 10^{-3}$	$p_1 = 10^{-2}$ $p_2 = 10^{-4}$	model dist. $\mu = 2.0727$ $\sigma^2 = 0.1882$	ET
quantiles	0.01	3.0401	2.9527	3.0819	3.1882
	0.001	3.3646	3.2568	3.4133	3.8214
	0.0001	3.6326	3.5154	3.6860	4.4547
interval	θ_1	4.2588	4.2588	$\times$	
	θ_2	7.5149	10.8842	$\times$	(critical
CVM dist		0.0870	0.1097	0.0795	value :
ET test	answer	accepted	accepted	accepted	0.1205)
	b_{inf}	2.7286	2.7742	2.7947	
	b_{sup}	5.5590	5.2760	5.6696	

Table 5: Bayesian regularization for the normal model, real data set

		predictive distribution $p_1 = 10^{-2}$ $p_2 = 10^{-3}$	$p_1 = 10^{-2}$ $p_2 = 10^{-4}$	model dist. $\mu = 0.7066$ $\sigma^2 = 0.0519$	ET
quantiles	0.01	3.0233	2.9539	3.4439	3.1882
	0.001	3.4578	3.3648	4.0986	3.8214
	0.0001	3.8693	3.7618	4.7299	4.4547
interval	θ_1	25.9642	25.9642	$\times$	
	θ_2	45.8149	66.3561	$\times$	(critical
CVM dist		0.1715	0.1996	0.0992	value :
ETtest	answer	accepted	accepted	accepted	0.1205)
	b_{inf}	2.6315	2.5449	3.1024	
	b_{sup}	5.4675	5.2372	6.8167	

Table 6: Bayesian regularization for the lognormal model, real data set

Finally, we also present the regularization procedure applied to the scale parameter of the Weibull model. Table 7 presents the results obtained with this procedure.
Here again, the quantiles obtained for both predictive distributions are close to those of the normal distribution. In both cases of overestimating (lognormal model) and underestimating (Weibull model) this normal tail, the expert opinions provide predictive quantiles closer to those of the normal ones. Finally, it appears clearly that according to the expert opinions, the normal model is the best one to fit this data set.

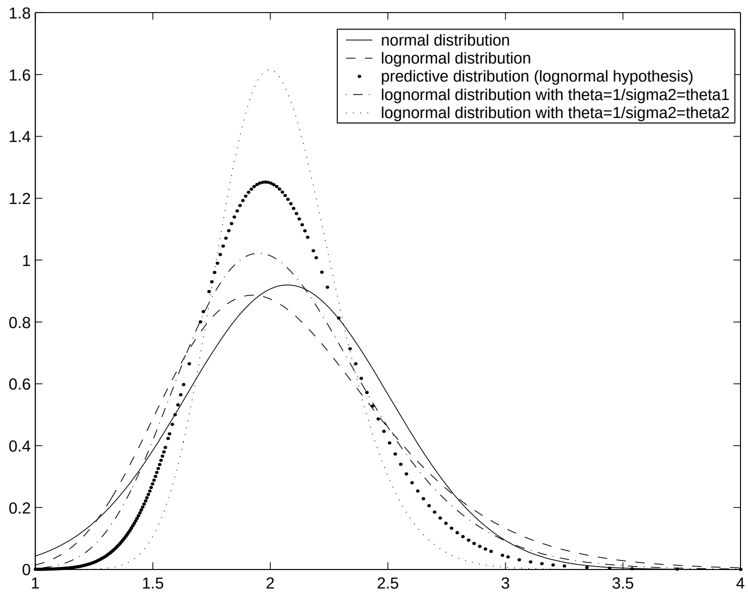

Figure 3: Real data set – density functions – comparing normal, lognormal and predictive (for lognormal model, $p_1 = 10^{-2}$ and $p_2 = 10^{-4}$) distributions

		predictive distribution		model dist.	ET
		$p_1 = 10^{-2}$ $p_2 = 10^{-3}$	$p_1 = 10^{-2}$ $p_2 = 10^{-4}$	$\alpha = 2.2382$ $\beta = 6.2877$	
quantiles	0.01	3.0827	2.9953	2.8535	3.1882
	0.001	3.2915	3.2032	3.0436	3.8214
	0.0001	3.4492	3.3620	3.1860	4.4547
interval	θ_1	0.0031	0.0031	$\times$	
	θ_2	0.0046	0.0061	$\times$	(critical
CVM dist		0.1696	0.0861	0.0761	value :
ET test	answer	accepted	accepted	accepted	0.1169)
	b_{inf}	2.9226	2.8702	2.7753	
	b_{sup}	5.3545	5.4482	5.0847	

Table 7: Bayesian regularization on scale parameter for the Weibull model, real data set

5 Final comments

In this work, we have presented a Bayesian regularization procedure designed to improve the fit of the tail, according to formatted expert opinions. Our numerical simulations show

that this is possible with only a small central deterioration. However, as expected with this Bayesian approach the correction is soft and for some models, even a strong regularization would not succeed in sufficiently improving the tail.
In practice, this procedure may be useful in different cases:

- When a centrally accepted distribution is rejected in the tail, the procedure proposes a better modelisation of the tail taking into account expert opinions. This was our initial motivation;
- When the distribution is accepted both centrally and in the tail, regularization is a Bayesian approach to build a better model;
- When several distributions are accepted both centrally and in the tail, this procedure provides indications to help select the best model in accordance with the opinions of the experts. This is what happened with our real data set.

We have also suggested how to use the information brought by the nonparametric estimation of the tail in this regularization procedure, when there is no expert opinion. However, getting closer to the Exponential Tail approximation can lead to more biased quantile estimates, as the ET estimate is itself biased.
Our procedure has been investigated for a set of usual families of distributions. This set has to be further enlarged. Distributions in the Fréchet domain of attraction would be of particular interest. As in the case of a prior on the shape parameter of the Weibull, this will probably lead to more computationally demanding procedures.

Acknowledgments

This work has been partially supported by the DER of Électricité de France. The authors wish to thank V. Durbec and B. Villain for stimulating and helpful discussions and for providing the real data set.

Appendix

We adopt the following distribution parametrisations:

$\mathcal{E}xp(\lambda)$: $\forall x \in R^+$ $\quad f(x) = \lambda \exp(-\lambda x)$

$\mathcal{G}amma(a,\, b)$: $\forall x \in R^{+*}$ $\quad f(x) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$

$\mathcal{N}(\mu,\, \sigma^2)$: $\forall x \in R$ $\quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$

$\mathcal{W}(\alpha,\, \beta)$: $\forall x \in R^{+*}$ $\quad f(x) = \frac{\beta}{\alpha}(\frac{x}{\alpha})^{\beta-1} \exp\left(-(\frac{x}{\alpha})^{\beta}\right)$

$\mathcal{B}eta(a,\, b,\, [\ell_m,\, \ell_M])$: $\forall x \in [\ell_m,\, \ell_M]$ $\quad f(x) = \dfrac{\Gamma(a+b)(x-\ell_m)^{a-1}(\ell_M - x)^{b-1}}{\Gamma(a)\Gamma(b)(\ell_M - \ell_M)^{a+b-1}}$

$GPD(\sigma,\, k)$: $k \neq 0$ $\quad \forall x \in R^{+*}$ if $k < 0$; $\forall x \in [0; \frac{\sigma}{k}[$ if $k > 0$ $\quad f(x) = \frac{1}{\sigma}(1 - \frac{kx}{\sigma})^{\frac{1}{k}-1}$

$k = 0$ $\quad \forall x \in R^{+*}$ $\quad f(x) = \frac{1}{\sigma} \exp(-\frac{x}{\sigma})$

References

Breiman, L., C. J. Stone, and C. Kooperberg (1990). "Robust confidence bounds for extreme upper quantiles". *Journal of Statistical Computation and Simulation 37*, 127–149.

Castillo, E. (1988). *Statistical Modeling and Decision Science.* New York: Academic Press.

Coles, S. G. and M. J. Dixon (1999). "Likelihood-based inference for extreme value models". *Extremes 2*, 5–23.

Coles, S. G. and E. A. Powell (1996). "Bayesian methods in extreme value modelling". *International Statistical Review 64*, 119–136.

Coles, S. G. and J. A. Tawn (1999). "A Bayesian analysis of extreme data rainfalls". *Applied Statistics 45*, 463–478.

D'Agostino, R. B. and M. A. Stephens (1986). *Goodness-of-fit Techniques.* Statistics : textbooks and monographs 68. New York and Basel: Marcel Dekker.

Diebolt, J., V. Durbec, M. A. El Aroui, and B. Villain (2000). "Estimation of extreme quantiles: empirical tools for method assessment and comparison". *International Journal of Reliability, Quality and Safety Engineering 7*(1), 75–94.

Diebolt, J., M. Garrido, and S. Girard (2001). Le test ET : test d'adéquation d'un modèle central à une queue de distribution. Technical Report RR-4170, Inria. http://www.inria.fr/rrrt/index.fr.html.

Ditlevsen, O. (1994). "Distribution arbitrariness in structural reliability". In S. Schuller and Yao (Eds.), *Structural Safety and Reliability*, pp. 1241–1247. Rotterdam: Balkema.

Embrechts, P., C. Klüpperberg, and T. Mikosch (1997). *Modelling Extremal Events for Insurance and Finance.* Stochastic Modelling and Applied Probability 33. Berlin: Springer-Verlag.

Garrido, M. and J. Diebolt (2000). "The ET test, a goodness of fit test to the tail distribution". In *Second International Conference on Mathematical Methods for Reliability*, Bordeaux - France.

Girard, S. and J. Diebolt (1999). "Consistency of the ET method and smooth variations". *Comptes Rendus de l'Académie des Sciences de Paris (Série I) 329*, 821–826.

Hahn, G. and W. Meeker (1982). "Pitfalls and practical considerations in product life analysis, part 1: Basic concepts and dangers of extrapolation". *Journal of Quality Technology 14*.

Hughey, R. (1991). "A survey and comparison of methods for estimating extreme right tail-area quantiles". *Communication in statistics - Theory Methods 20*(4), 1463–1496.

Moon, Y., U. Lall, and K. Bosworth (1993). "Comparison of tail probability estimators for flood frequency analysis". *Journal of Hydrology 151*(2-4), 343–363.

Pickands (1975). "Statistical inference using extreme order statistics". *Annals of Statistics 3*, 119–131.

Contents

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399